Cognition 110 (2009) 254–259

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Brief article

# Phonetic recalibration only occurs in speech mode

Jean Vroomen *, Martijn Baart

*Department of Psychology, Tilburg University, P.O. Box 90153, Warandelaan 2, 5000 LE, Tilburg, The Netherlands*

ARTICLE INFO

*Article history:*
Received 28 May 2008
Revised 12 September 2008
Accepted 24 October 2008

*Keywords:*
Recalibration of phonetic categories
Lipread speech
Phonetic learning
Selective speech adaptation
Phoneme perception

ABSTRACT

Upon hearing an ambiguous speech sound dubbed onto lipread speech, listeners adjust their phonetic categories in accordance with the lipread information (recalibration) that tells what the phoneme should be. Here we used sine wave speech (SWS) to show that this tuning effect occurs if the SWS sounds are perceived as speech, but not if the sounds are perceived as non-speech. In contrast, selective speech adaptation occurred irrespective of whether listeners were in speech or non-speech mode. These results provide new evidence for the distinction between a speech and non-speech processing mode, and they demonstrate that different mechanisms underlie recalibration and selective speech adaptation.



## 1. Introduction

A critical question about speech is whether specialized processors are responsible for the coding of the acoustic signal in phonetic segments (Liberman & Mattingly, 1985) or whether speech is perceived as all other sounds (Massaro, 1987). A clear demonstration of the existence of a speech versus non-speech mode was provided by Remez, Rubin, Pisoni, and Carrell (1981) using sine wave speech (SWS). In SWS, the natural richness of the auditory signal is reduced to a few sinusoids (usually three) that follow the centre frequency and the amplitude of the first three formants. These stimuli sound highly artificial, and most naïve subjects perceive them as 'non-speech' sounds like whistles or sounds from a science fiction movie. Typically, though, once subjects are told that these sounds are actually derived from speech, they cannot switch back to a non-speech mode again and continue to hear the sounds as speech. Functional brain imaging studies have provided converging evidence as for listeners in speech mode there is stronger activity in the left superior temporal sulcus than for listeners in non-speech mode (Möttönen et al., 2006). Moreover, if SWS sounds are combined with lipread speech, then naïve subjects in non-speech mode show no or only negligible intersensory integration (lipread information biasing speech sound identification), while subjects who learned to perceive the same auditory stimuli as speech do integrate the auditory and visual stimuli in a similar manner as natural speech (Tuomainen, Andersen, Tiippana, & Sams, 2005).

Previous studies demonstrating the speech/non-speech mode distinction had to rely on the immediate subjective report that the SWS stimuli were actually perceived as speech or non-speech. Here, we demonstrate that there are also indirect effects using two distinct phenomena that we hypothesized to be differently sensitive as to whether perceivers were in speech or non-speech mode, namely recalibration of phonetic categories and selective speech adaptation. Recalibration of phonetic categories is a tuning effect that occurs when a phonetically ambiguous speech sound is combined with lipread speech. While being exposed to such an audiovisual stimulus, participants adjust the phoneme boundary and learn to categorize the initially ambiguous speech sound in accordance with the simultaneously presented lipread speech. This can be demonstrated in a subsequent auditory-only test where listeners identify the ambiguous sound. For example, if an ambiguous sound

* Corresponding author. Tel.: +31 13 4662394; fax: +31 13 4662370.
*E-mail address:* j.vroomen@uvt.nl (J. Vroomen).

0010-0277/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2008.10.015

halfway between /b/ and /d/ is dubbed onto lipread /b/, then participants are more likely to categorize the ambiguous sound as /b/. Presumably, recalibration is induced by the deviance between the heard and lipread information that the brain tries to minimize by shifting the phoneme boundary (Bertelson, Vroomen, & de Gelder, 2003; van Linden & Vroomen, 2007; van Linden & Vroomen, in press; Vroomen, van Linden, de Gelder, & Bertelson, 2007; Vroomen, van Linden, Keetels, de Gelder, & Bertelson, 2004).

Selective speech adaptation, first demonstrated by Eimas and Corbit (1973), is different from recalibration in that it does not depend on a conflict between two information sources, but rather depends on the repeated presentation of a particular speech sound by itself that causes a reduction in the frequency with which that token is reported in subsequent identification trials. Since its introduction, many questions have been raised about the nature underlying this effect. Originally, it was thought to reflect a fatigue of some hypothetical 'linguistic feature detectors', but others argued that it reflects a shift in criterion (Diehl, Elman, & McCusker, 1978), or a combination of both (Samuel, 1986). Still others (e.g. Ganong, 1978) showed that the size of selective speech adaptation depends upon the degree of spectral overlap between the adapter and test sound, and that most, if not all of the effect is auditory rather than phonetic. A similar conclusion was reached by Roberts and Summerfield (1981). They exposed listeners to audiovisual congruent (auditory /b/ with lipread /b/) or incongruent adapter stimuli (auditory /b/ with lipread /g/) and obtained similar aftereffects, despite that the adapters were perceived differently. Selective adaptation thus mainly depends on the acoustic nature of the adapter, and not the lipread component or the phonetic percept (see also Saldaña and Rosenblum, 1994).

Here, we examined whether recalibration and selective speech adaptation occurs with SWS stimuli, and whether the effects would differ for listeners in speech versus non-speech mode. We hypothesized that lipread-induced recalibration occurs if, and only if perceivers are in speech mode but not in non-speech mode because in non-speech mode there is no intersensory integration (Tuomainen et al., 2005) and hence no phonetic conflict between sight and sound that would induce recalibration. We thus assumed that recalibration occurs to the extent that conflicting information sources are referring to the same event. If listeners are in speech mode, heard and lipread inputs are combined into a single phonetic representation, but not so if listeners are not under the impression that the auditory and visual signals refer to separate events. Selective adaptation, though, may occur for listeners in speech and non-speech mode, assuming that this phenomenon depends on some low-level acoustic factor and not the phonetic interpretation of the sound (Roberts & Summerfield, 1981).

To test these hypotheses, we created an SWS continuum between /omso/ and /onso/. Participants were trained to categorize the two auditory endpoints of this continuum as /omso/ or /onso/ for the speech group, or as '1' or '2' for the non-speech group. Once participants reliably discriminated the two sounds, they were exposed to audiovisual adapter stimuli intended to induce recalibration or selective speech adaptation and then tested. To induce recalibration, we used audiovisual adapters containing the most ambiguous SWS token of the continuum halfway between /omso/ and /onso/ (henceforth /A?/ for 'Auditory ambiguous') dubbed onto a video recording of the speaker articulating /omso/ or /onso/ ($A?V_{omso}$ and $A?V_{onso}$). Following a short exposure phase, auditory-only test trials were given in which participants identified the SWS tokens from the middle of the continuum. For participants in speech mode, we expected the ambiguous tokens to be labelled in accordance with the previously seen lipread adapter, so more /onso/ responses after exposure to $A?V_{onso}$ than $A?V_{omso}$. No such difference was expected for the non-speech group, because lipread speech should not affect the auditory tokens if they are labelled as non-speech (Tuomainen et al., 2005).

To induce selective adaptation, we used audiovisual adapters containing the endpoint tokens of the /omso/–/onso/ continuum, and dubbed these onto congruent video recordings of the speaker. Participants were thus exposed to $A_{omso}V_{omso}$ and $A_{onso}V_{onso}$. Due to the non-ambiguous acoustic nature of the sound, we expected to observe contrastive aftereffects irrespective of whether participants were in speech or non-speech mode, so more /onso/- or '2'-responses after exposure to $A_{omso}V_{omso}$ and more /omso/ or '2'-responses after $A_{onso}V_{onso}$.

## 2. Method

### 2.1. Participants

Twenty-four native speakers of Dutch (first-year students) participated. Half of them were trained in speech mode, the other half in non-speech mode.

### 2.2. Stimuli

Stimulus creation started from the original recording of natural /omso/ and /onso/ tokens previously used by Tuomainen et al. (2005). Using the Praat-programme (Boersma & Weenink, 2005), a seven-point continuum between /omso/ and /onso/ was created by changing the second (F2) and third (F3) formants in equal steps. The steady state value of the F2 in the initial vowel was 780 Hz and lasted 140 ms for both endpoints. The transition of the F2 in the nasal was 50 ms, and its offset frequency varied from 1800 Hz for the /onso/-endpoint to 680 Hz for the /omso/-endpoint in equal Mel steps. The F3 had a steady state value of 2500 Hz in the vowel, and the offset frequency of the transition varied from 2500 Hz for the /onso/-endpoint to 2250 Hz for the /omso/-end point. This resulted in a natural sounding seven-point /omso/–/onso/ continuum. Pilot tests showed ($N = 16$) that the middle (fourth) stimulus was also the most ambiguous one (Fig. 1).

The tokens of the thus created continuum were transformed into SWS sounds using a script from C. Darwin available on the internet (http://www.biols.susx.ac.uk/home/Chris_Darwin/Praatscripts/SWS). Three-tone SWS stimuli were created with time varying sine waves for the three lowest formants (Fig. 2). These SWS stimuli where then dubbed onto the video recording of the speaker

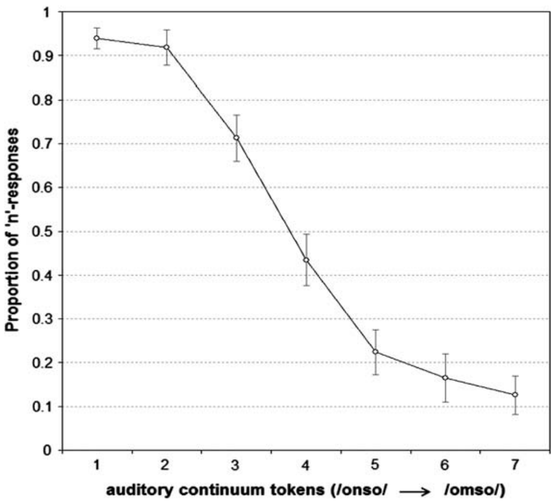


**Fig. 1.** Mean proportion of /onso/ responses of the original synthetic continuum. Error bars represent one standard error of the mean.

(29.97 frames per s., 22 × 15 cm) articulating either /omso/ or /onso/, preserving the natural timing between the audio and video. This resulted in four audiovisual adapter stimuli: $A?V_{omso}$ and $A?V_{onso}$ (to induce recalibration) and $A_{omso}V_{omso}$ and $A_{onso}V_{onso}$ (to induce selective speech adaptation). The sound level of the stimuli peaked at 79 dBa when measured at ear level.

To ensure that participants where looking at the screen during adaptation, participants had to detect a small white dot that appeared for 100 ms on the upper lip of the speaker. Participants had to press a special key upon appearance of such an occasional catch trial.

### 2.3. Procedure and design

Participants were tested individually in a sound attenuated and dimly lit room at 70 cm distance from a 17 inch CRT monitor. They were first acquainted with the SWS tokens and learned to categorize the endpoints as /omso/ and /onso/ for the speech group, or as '1' and '2' for the non-speech group. The two endpoints were delivered 48 times in pseudorandom order with immediate feedback. Participants continued training without corrective feedback until a learning criterion was met (12 consecutively correct answers). Two participants (one in speech mode, the other in non-speech mode) failed to meet this criterion after a predetermined time limit and were replaced. The learning criterion was reached after 33.0 trials for the speech group, and 24.3 trials for the non-speech group; $t(22) = .86$, $p = .40$. From the start, both groups were thus equally good in discriminating the auditory SWS endpoints.

### 2.4. Adapter-test blocks

Similar procedures were used as in Bertelson et al. (2003, Experiment 2). Participants were repeatedly exposed to short blocks of audiovisual adapter stimuli immediately followed by auditory-only test trials. Each adapter block contained eight consecutively presented adapter stimuli (either $A?V_{omso}$, $A?V_{onso}$, $A_{omso}V_{omso}$, or $A_{onso}V_{onso}$, ISI = 425 ms) followed by six test trials. In the test, the most ambiguous SWS token (A?) and the more /onso/-like (A? − 1) and /omso/-like stimulus (A? + 1) of the continuum were presented twice. Participants pressed a designated key upon perceiving /omso/ (or '1') or /onso/ (or '2'). Participants were exposed to eight blocks of each adapter (32 adapter-test blocks in total), all presented in pseudorandom order.

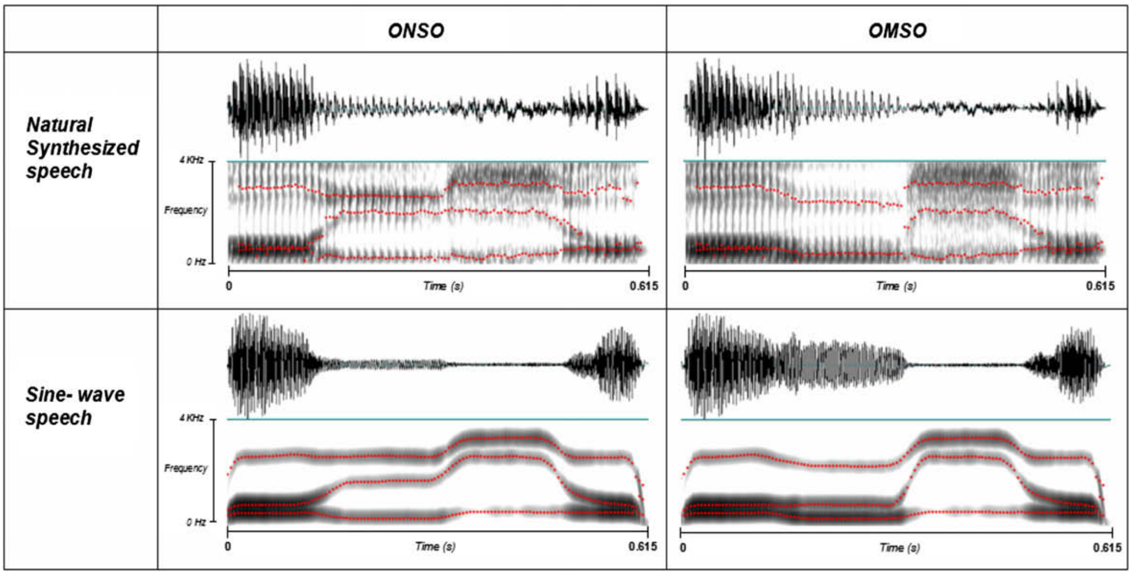


**Fig. 2.** Waveforms and corresponding spectograms of the endpoints of the synthesized continuum and their sine wave replicas. Formants (F1, F2 and F3) are represented by dotted lines.

### 2.5. Goodness rating of adapters

In the final part, participants rated the auditory quality of the audiovisual adapter stimuli. Each adapter was presented six times in pseudorandom order and participants rated the goodness of the sound on a seven-point Likert scale with '1' for a clear /omso/ or '1', and '7' for a clear /onso/ or '2'. Finally, participants in the non-speech group were asked whether they had noticed that the SWS stimuli originated from actual speech. Three reported to have heard spoken syllables (though not /omso/ and /onso/) and were replaced by others.

## 3. Results

Performance on catch trials was almost flawless (99% correct for the speech group and 96% correct for the non-speech group) indicating that participants were indeed looking at the video during exposure to lipread speech.

### 3.1. Goodness ratings of adapters

The goodness ratings of the audiovisual adapters were analyzed first to ensure that they were perceived as intended. As in Tuomainen et al. (2005), lipreading had a strong impact on the ambiguous SWS sound if and only if the sound was perceived as speech, but not if perceived as non-speech (see Table 1). In the 2 (speech/non-speech mode) × 2 (ambiguity of adapter sound) × 2 (lipread /omso/ or /onso/) overall ANOVA, the critical interaction between mode, ambiguity of adapter sound, and lipread adapter was highly significant, $F(1,22) = 16.34$, $p < .002$ ($\eta^2 = .43$). A separate ANOVA for adapter stimuli with ambiguous sounds ($A?V_{omso}$, $A?V_{onso}$) showed the main effect of lipreading, $F(1,22) = 19.77$, $p < .001$ ($\eta^2 = .47$), interacted with speech mode, $F(1,22) = 19.80$, $p < .001$ ($\eta^2 = .47$). Separate *t*-test confirmed that lipreading affected the quality of the ambiguous sound if listeners were speech mode (a 2.79 bias, $t(11) = 4.98$, $p < 0.001$), but not so if listeners were in non-speech mode (0.00 bias, testing unneeded). The ANOVA for adapter stimuli with non-ambiguous sounds ($A_{omso}V_{omso}$, or $A_{onso}V_{onso}$) showed there was a significant stimulus effect, $F(1,22) = 487.68$, $p < .001$ ($\eta^2 = .96$), but no interaction with speech mode ($F < 1$). Non-ambiguous sounds were thus equally distinct for both groups.

### 3.2. Test trials

Performance on test trials following exposure to the different adapters is presented in Fig. 3. We also computed aftereffects as in previous studies (Bertelson et al., 2003) by subtracting the proportion of /onso/ responses following exposure to /omso/ from /onso/ pooling over the three test tokens (see Table 2). As is clearly visible, for the speech group there was recalibration and selective speech adaptation, while for the non-speech group there was only selective adaptation with no sign of recalibration. In the 2 (speech/non-speech mode) × 2 (ambiguity of adapter sound) × 2 (lipread /omso/ or /onso/) × 3 (Auditory test token) overall ANOVA, there was a main effect of ambiguity of the adapter sound $F(1,22) = 5.80$, $p < .025$ ($\eta^2 = .21$), because there were more /onso/ responses after exposure to non-ambiguous adapter sounds, and a main effect of auditory test token, $F(2,44) = 63.62$, $p < .001$ ($\eta^2 = .74$), because there were more /onso/- (or '2'-) responses for sounds from the /onso/- than /omso/-side of the continuum. The interaction between the ambiguity of the adapter sound and lipread speech was significant, $F(1,22) = 31.57$, $p < .001$ ($\eta^2 = .59$), because there were more /onso/ responses after exposure to $A?V_{onso}$ than $A?V_{omso,}$ (i.e., recalibration), but less /onso/ responses following exposure to $A_{onso}V_{onso}$ than $A_{omso}V_{omso}$ (i.e., selective speech adaptation). Most important, the size of this effect differed for the speech and non-speech group as reflected in a significant second-order interaction, $F(1,22) = 5.14$, $p < .034$ ($\eta^2 = .19$). Separate *t*-tests confirmed that for the speech group, there were 14% more /onso/ responses after exposure to $A?V_{onso}$ than $A?V_{omso}$, $t(11) = 3.96$, $p < .002$ (one-sided, as there was a clear prediction), while there were 19% less /onso/ responses after exposure to $A_{onso}V_{onso}$ than $A_{omso}V_{omso}$, $t(11) = 2.40$, $p < .036$. For the non-speech group, there was no difference (0%) between $A?V_{onso}$ and $A?V_{omso}$,; $t(11) = .\ 17$, $p < .87$, whereas there were 13% less /onso/ responses after exposure to $A_{onso}V_{onso}$ than $A_{omso}V_{omso}$, $t(11) = 4.59$, $p < .001$ (selective speech adaptation).

**Table 1**
Goodness ratings of the audiovisual adapters. Lipread bias was calculated by subtracting /omso/ ratings from /onso/.

| Mode | Exposure sound | Lipread adapter | | Lipread bias |
|---|---|---|---|---|
| | | /onso/ | /omso/ | |
| Speech mode | Ambiguous | 5.27 | 2.48 | 2.79 |
| | Non-ambiguous | 6.50 | 1.69 | 4.81 |
| Non-speech mode | Ambiguous | 4.19 | 4.19 | 0.00 |
| | Non-ambiguous | 6.37 | 1.74 | 4.63 |

## 4. Discussion

The present results clearly demonstrate that recalibration of phonetic categories and selective speech adaptation can be obtained with sine wave replicas. Moreover, the use of SWS allowed us to observe a remarkable dissociation between these two phenomena: recalibration was observed only when listeners were in speech mode, whereas selective adaptation occurred for listeners in speech and non-speech mode. Previous studies already demonstrated that these two phenomena not only differ in the direction of their aftereffect, but also in the speed with which they build-up and dissipate (recalibration builds up fast and peaks early, selective adaptation builds up slowly and increases with prolonged exposure (Vroomen et al., 2004, 2007)). Together, these dissociations therefore provide strong evidence that there are distinct mechanisms underlying recalibration and selective adaptation.

Our findings on selective speech adaption fit well with previous reports showing that low-level mechanism are mainly responsible for the effect to occur. For example, Roberts and Summerfield (1981) demonstrated that adaptation was induced by the auditory component, whereas the phonetic label attached to the adapting stimulus had

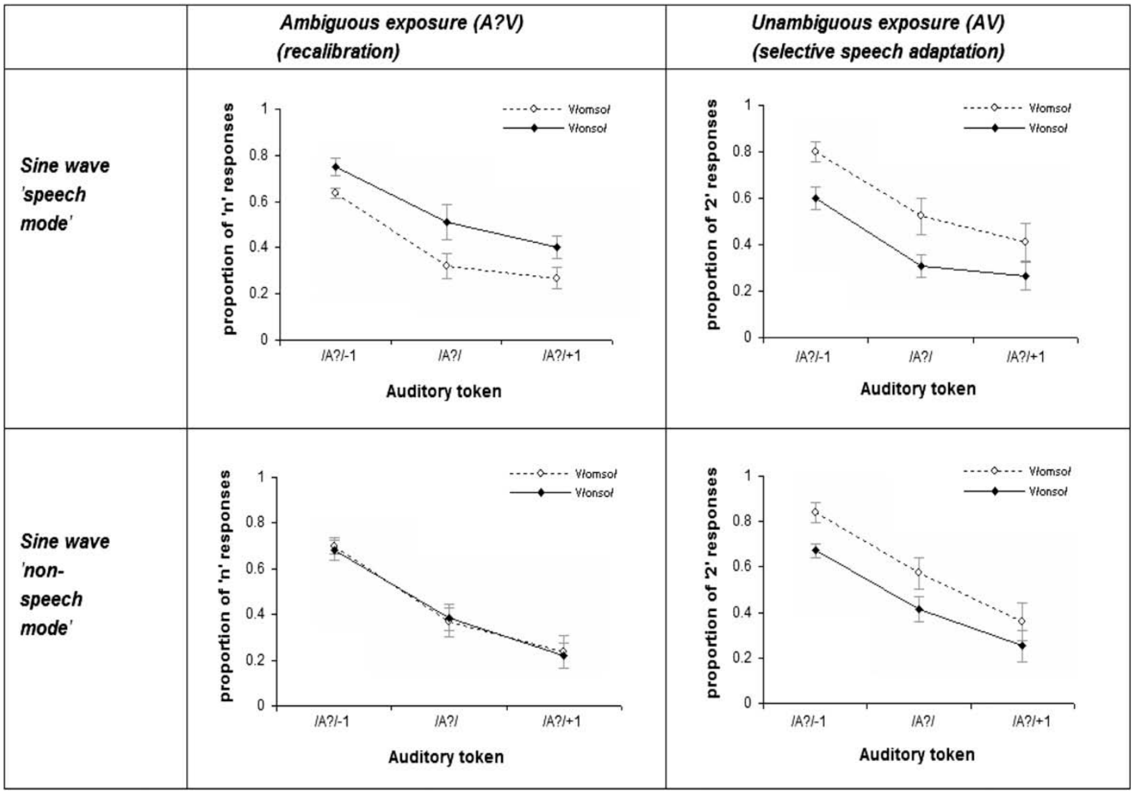


**Fig. 3.** Mean proportion of /onso/ responses as a function of the auditory test tokens of the continuum after exposure to auditory ambiguous adapters $A?V_{onso}$ and $A?V_{omso}$ (left panels), and auditory non-ambiguous adapters $A_{onso}V_{onso}$ and $A_{omso}V_{omso}$ (right panels). The upper panels show performance of the speech group, the lower panels of the non-speech group. Error bars represent one standard error of the mean.

**Table 2**
Mean proportion of 'onso'- or '2'-responses and the corresponding aftereffect after exposure to audiovisual adapters with ambiguous and non-ambiguous sounds.

| Mode | Exposure sound | Lipread adapter | | Aftereffect |
|---|---|---|---|---|
| | | /onso/ | /omso/ | |
| Speech mode | Ambiguous | .55 | .41 | .14* |
| | Non-ambiguous | .39 | .58 | –.19* |
| Non-speech mode | Ambiguous | .43 | .43 | .00 |
| | Non-ambiguous | .46 | .59 | –.13* |

* Significance at $p < .05$.

no effect. Here we also observed that equal amounts of selective adaptation were obtained for listeners in speech or non-speech mode. This again suggests that it is the acoustic and non-ambiguous nature of the adapter that causes selective adaptation, while the more high-level interpretation of the stimulus has little or no effect. In that sense, adaptation is also similar to other forms of perception like color, curvature (Gibson, 1933) or motion (Anstis, 1986, chap. 16) where aftereffects mainly depend on the non-ambiguous visual nature of the adapting stimulus.

In stark contrast with selective adaptation, recalibration appeared to be speech-specific. The notion underlying recalibration is that reliable information from one source disambiguates unreliable information from another source. Here, it was lipread speech that provided reliable information about how to interpret an ambiguous 'm/n' sound. Presumably, during exposure there is a conflict between the heard and lipread information that is resolved by shifting the phoneme boundary so that the ambiguous sound matches the lipread information. This shift occurs quite fast (Vroomen et al., 2007) and it lasts for some time so that it is observable as an aftereffect. It seems only logical that recalibration occurs to the extent that the conflicting information sources are referring to the same distal event, here whether the speaker said /m/ or /n/. For listeners in speech mode, both inputs were indeed combined into a single phonetic presentation as observable in a direct bias effect on the goodness rating of the sound. Listeners in non-speech mode, though, were not under the impression that the auditory and visual signal referred to the same event, and the two information streams were therefore treated as separate. Listeners labelling the SWS sounds as '1' or '2' thus made no connection with the segmental content of the simultaneously presented lipread information, and there was therefore also no effect of lipreading on the goodness ratings of the SWS sound if perceived as non-speech.

The use of the SWS stimuli to induce recalibration and selective speech adaptation may also provide new opportunities to explore the nature of these phenomena. Eisner and McQueen (2005) reported that recalibration for the fricatives (/s/–/f/) did not generalize to a novel speaker. Similarly, Kraljic and Samuel (2005) tested the fricatives (/s/–/ʃ/) and found that tuning did not generalize across speakers. When a male voice was heard during the exposure phase, at test recalibration was reliable for male-produced tokens but not for female-produced tokens. Kraljic and Samuel (2006) also tested stop consonants (/d/–/t/) and here they did observe that recalibration generalized to a novel speaker. They argued that the patterns of generalization may be due to the acoustic similarity among the different exposure and test tokens. On this view, recalibration generalizes to acoustically similar sounds, but not to acoustically dissimilar sounds (see also Mirman, McClelland, & Holt, 2006). It remains for future studies to explore whether there is generalization from SWS sounds to natural speech and vice versa, and whether the same holds for selective speech adaptation. If it is indeed the acoustic similarity across tokens that determines whether recalibration will generalize, one may find that there is no generalization from SWS tokens to natural speech, while there is generalization for selective speech adaptation.

## References

Anstis, S. (1986). Motion perception in the frontal plane: Sensory aspects. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.). *Handbook of perception and human performance* (Vol. 1). New York: Wiley.

Bertelson, P., Vroomen, J., & de Gelder, B. (2003). Visual recalibration of auditory speech identification: A McGurk aftereffect. *Psychological Science, 14*(6), 592–597.

Boersma, P., & Weenink, K. (2005). Praat: Doing phonetics by computer (Version 4.3.14).

Diehl, R. L., Elman, J. L., & McCusker, S. B. (1978). Contrast effects on stop consonant identification. *Journal of Experimental Psychology: Human Perception and Performance, 4*(4), 599–609.

Eimas, P. D., & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology, 4*, 99–109.

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception and Psychophysics, 67*(2), 224–238.

Ganong, W. F. (1978). The selective adaptation effects of burst-cued stops. *Perception and Psychophysics, 24*(1), 71–83.

Gibson, J. J. (1933). Adaptation, after-effects and contrast in the perception of curved lines. *Journal of Experimental Psychology, 18*, 1–31.

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology, 51*(2), 141–178.

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin and Review, 13*(2), 262–268.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*(1), 1–36.

Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc..

Mirman, D., McClelland, J. L., & Holt, L. L. (2006). An interactive Hebbian account of lexically guided tuning of speech perception. *Psychonomic Bulletin and Review, 13*(6), 958–965.

Möttönen, R., Calvert, G. A., Jääskeläinen, I. P., Matthews, P. M., Thesen, T., Tuomainen, J., et al (2006). Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. *Neuroimage, 30*(2), 563–569.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science, 212*(4497), 947–949.

Roberts, M., & Summerfield, Q. (1981). Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory. *Perception and Psychophysics, 30*(4), 309–314.

Saldaña, H. M., & Rosenblum, L. D. (1994). Selective adaptation in speech perception using a compelling audiovisual adaptor. *Journal of the Acoustical Society of America, 95*(6), 3658–3661.

Samuel, A. G. (1986). Red herring detectors and speech perception: In defense of selective adaptation. *Cognitive Psychology, 18*(4), 452–499.

Tuomainen, J., Andersen, T. S., Tiippana, K., & Sams, M. (2005). Audio–visual speech perception is special. *Cognition, 96*(1), B13–B22.

van Linden, S., & Vroomen, J. (2007). Recalibration of phonetic categories by lipread speech versus lexical information. *Journal of Experimental Psychology: Human Perception and Performance, 33*, 1483–1494.

van Linden, S., & Vroomen, J. (in press). Audiovisual speech recalibration in children. *Journal of Child Language*.

Vroomen, J., van Linden, S., de Gelder, B., & Bertelson, P. (2007). Visual recalibration and selective adaptation in auditory-visual speech perception: Contrasting build-up courses. *Neuropsychologia, 45*, 572–577.

Vroomen, J., van Linden, S., Keetels, M., de Gelder, W., & Bertelson, P. (2004). Selective adaptation and recalibration of auditory speech by lipread information: Dissipation. *Speech Communication, 44*(1–4), 55–61.